DOCUMENT RESUME

ED 400 320 TM 025 716

AUTHOR Webster, William J.; And Others
TITLE The Applicability of Selected Regression and Hierarchical Linear Models to the Estimation of School and Teacher Effects.
PUB DATE Apr 96
NOTE 35p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Bayesian Statistics; Correlation; Effective Schools Research; *Estimation (Mathematics); *Least Squares Statistics; Mathematical Models; Public Schools; *Regression (Statistics); *School Effectiveness; Statistical Bias; *Teacher Effectiveness
IDENTIFIERS *Dallas Independent School District TX; *Hierarchical Linear Modeling

ABSTRACT

Five issues relative to the use of different Ordinary Least Squares (OLS) and Hierarchical Linear Modeling (HLM) models to identify effective schools and teachers were examined using data from all students in the Dallas (Texas) public schools in grade 3 in 1994 and grade 4 in 1995. OLS models using first- and second-order interactions produced results that were very close to those produced by two-level HLM models at the school level and two- and three-level HLM models at the teacher level. Most OLS regression and HLM models used in this study accounted for more than 70% of the variance in student achievement in reading and mathematics. Results produced by all the models were extremely consistent, and correlations produced by the various models were all generally above 0.90. Correlations of results with important school, teacher, and student-level contextual variables were negligible for all models, meaning that the various models produced results that were free from bias relative to important contextual variables. Correlations of results with prescore characteristics were negligible for all models, meaning that the various models produced results that were free from bias relative to the level of pretest scores. Taking all results into consideration, it is recommended that a two-level HLM model (student-school) be implemented to determine school effect, and that the empirical Bayes residuals from the model be adjusted with an adjustment for shrinkage to form the basis for estimates of teacher effect. Appropriate formulas for this task are included. (Contains 34 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The Applicability of Selected Regression and Hierarchical Linear Models to The Estimation of School and Teacher Effects[1]

ED 400 320

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

William J. Webster, Robert L. Mendro,
Timothy H. Orsak, and Dash Weerasinghe
Dallas Public Schools

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
WILLIAM J. WEBSTER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Abstract

This study has examined five issues relative to the use of different OLS regression and HLM models in identifying effective schools and teachers. First, OLS regression models using first and second order interactions produced results that were very close to those produced by two-level HLM models at the school level and two and three-level HLM models at the teacher level. Second, most OLS regression and HLM models used in this study accounted for more than seventy percent of the variance in student achievement in reading and mathematics. Generally, as more information was included in the equations more variance was accounted for. Third, the results produced by all of the models were extremely consistent. The correlations among results produced by the various models were all generally above .90. Fourth, correlations of results with important school, teacher, and student level contextual variables were negligible for all models meaning that the various models produced results that were free from bias relative to important school, student, and classroom level contextual variables. Fifth, correlations of results with pre-score characteristics were negligible for all models meaning that the various models produced results that were free from bias relative to level of pretest scores. Taking all results into consideration, the recommended solution was to implement a two-level HLM model (student-school) to determine school effect and then to adjust the empirical bayes residuals from that model with an adjustment for shrinkage to form the basis for the estimates of teacher effect. This paper concludes with the appropriate formulas for accomplishing this.

The need for instructional improvement in the Dallas Public Schools, like most urban districts, had been thoroughly documented over a period of twenty years. After a period of rapid achievement growth in the early and mid 1980's, student achievement in the Dallas schools had leveled out. In 1990, responding to this need, the District's Board of Education appointed a citizen's task force, the Commission for Educational Excellence, to formulate recommendations to accelerate the needed improvement. After a year of community hearings and extensive study, the Commission recommended a six point plan for massive educational reform. At the heart of the Commission's recommendations was an accountability system that fairly and accurately evaluated schools and teachers on their contributions to accelerating student growth in a number of important and valued outcomes of schooling. This was coupled with a movement to give schools more decision-making authority about personnel, curriculum, and most other aspects of schooling. In exchange for this authority, school staffs were to be held accountable for their

TM025716

[1] Paper presented at the annual meeting of the National Council on Measurement in Education, April 8-12, 1996.

BEST COPY AVAILABLE

actions. As part of this recommendation, $2.4 million was set aside as an incentive award to reward effective schools and their professional and support staffs.

It then became the task of the District's Research, Planning, and Evaluation Department to develop, pilot test, and implement an evaluation system to accomplish the goals of the Commission. The first step in accomplishing this task was the appointment of an Accountability Task Force to oversee the process. This task force, consisting of teachers, principals, parents, members of the business community, and central office administrators, was charged with the responsibility of advising the General Superintendent concerning the implementation of a performance incentive plan, working with the administration to ensure the validity of the selection procedure and subsequent results of the incentive plan, and serving as a review committee to examine any issues raised by personnel concerning questions of equity and fairness of the procedures. During a year of exhaustive deliberations, a number of requirements for the methodology associated with this plan were developed. Among these were:

1. It must be value-added.

2. It must include multiple outcome variables.[2]

3. Schools must only be held accountable for students who have been exposed to their instructional program (continuously enrolled students).

4. It must be fair. Schools must derive no particular advantage by starting with high-scoring or low-scoring students, minority or white students, high or low socioeconomic level students, or limited English proficient or non-limited English proficient students. In addition such factors as student mobility, school overcrowding, and staffing patterns over which the schools have no control must be taken into consideration.

5. It must be based on cohorts of students, not cross-sectional data.

Within the five aforementioned parameters, a number of statistical models are possible. This study examines alternative methodologies for determining school effect then extends these studies to the determination of teacher effect. These models are designed to isolate the effect of a

---

[2] Performance indicators for 1995-96 include *Iowa Texas of Basic Skills* and *Test of Achievement and Proficiency* reading and mathematics, grades 1-9, *Spanish Assessment of Basic Education*, grades 1-6;*Texas Assessment of Academic Skills* reading and mathematics, grades 3-8 and 10; writing, grades 4,8, and 10; science and social studies, grade 4 and 8; *Texas Assessment of Academic Skills*, Spanish version grades 3 and 4; 72 standardized final examinations in language, mathematics, social studies, science, ESOL, reading, and world languages, grades 9-12; promotion rate, grades 1-8; student attendance, grades 1-12; graduation rate, grades 9-12; *Scholastic Aptitude Test* percent tested and scores, grades 9-12; dropout rate, grades 7-12; student enrollment in prehonors/honors courses, grades 7-12; student enrollment in advanced diploma plans, grades 9-12; students enrolled in advanced placement courses, grades 11-12; *Preliminary Scholastic Aptitude Test* percent tested and scores, grades 9-12; and percent passing Advance Placement Exams, grades 11-12. The system is run with only continuously enrolled students and includes staff attendance incentive, minimum percent eligible tested requirements, and requirements that at least one-half of a school's cohorts must outgrow the national norm group on the *ITBS* and TAP in reading and mathematics.

school's or teacher's practices on important student outcomes. The school effect can be conceptualized as the difference between a given student's performance in a particular school and the performance that would have been expected if that student had attended a school with similar context but with practice of average effectiveness. The teacher effect can be conceptualized similarly at the teacher level.

Background

Interest in performance-based or outcomes-based teacher evaluation dates all the way back to fifteenth-century Italy where a teacher master's salary was dependent upon his or her students' performance. Despite long-term interest, progress in actually linking student outcomes to school and teacher performance has been very limited.

State Departments of Education have taken a leadership role in attempting to accomplish this at the school and district level. Forty-six of fifty states have accountability systems that feature some type of assessment. Twenty-seven of these systems feature reports at the school, district, and state level; three feature school level reports only; six feature reports at both the school and district level; seven feature reports at the district and state level; two feature reports at the state level only; and one is currently under development (Council of Chief State School Officers, 1995). When one reviews these systems, it is obvious that their designers are not familiar with the literature on value-added systems since only two states, South Carolina (May, 1990) and Tennessee (Sanders and Horn, 1995) have used appropriate value-added statistical methodology in implementing such systems. Most of the rest tend to evaluate students, not schools or districts, and generally cause more harm than good with systematic misinformation about the contributions of schools and districts to student academic accomplishments. By comparing schools on the basis of unadjusted student outcomes, state reports are often systematically biased against schools with population demographics that differ from the norm, a fact that was graphically illustrated by Jaeger (1992). In attempting to eliminate this bias, a number of states have gone to non-statistical grouping techniques, an approach that has serious limitations when there is consistent one-directional variance on the grouping characteristics within groups.

Investigators throughout the world have conducted and reported numerous studies aimed at identifying effective schools as well as estimating the magnitude and stability of school contributions to student outcomes. Good and Brophy (1986) provide an excellent review of this work. Researchers have been working for a number of years on appropriate methodology for adjusting for the effects of student and school demographic variables in estimating school effects. One approach has been to regress school mean outcome measures on school means of one or more background variables. This approach is only adequate to the extent that there is not much within school variance, that is, the school impacts all students similarly. Mendro and Webster, (1993) demonstrated that this is not the case and that using school level models to attempt to estimate school effects, while better than the common practice of reporting unadjusted test scores, produces extremely unstable estimates of school effects.

Another approach, one that has received generally widespread acceptance among educational researchers, involves the aggregation of residuals from student-level regression models (Aiken and West, 1991; Bano, 1985; Felter and Carlson, 1985; Kirst, 1986; Klitgard and Hall, 1973; McKenzie, 1983; Millman,1981; Saka, 1984; Webster and Olson, 1988; Webster, Mendro, and Almaguer, 1994). These techniques can incorporate a large number of input, process, and outcome variables into an equation and determine the average deviation from the predicted student outcome values for each school. Schools are then ranked on the average deviation. Some advantages of multiple regression analysis over other statistical techniques for this application include its relative simplicity of application and interpretation, its robustness, and the fact that general methods of structuring complex regression equations to include combinations of categorical and continuous variables and their interactions are relatively straightforward (Aiken and West, 1991; Cohen, 1968; Cohen and Cohen, 1983; Darlington, 1990).

Finally, hierarchical linear modeling (HLM) provides estimates of linear equations that explain outcomes for group members as a function of the characteristics of the group as well as the characteristics of the members. Because HLM involves the prediction of outcomes of members who are nested within groups which in turn may be nested in larger groups, the technique should be well suited for use in education. The nested structure of students within classrooms and classrooms within schools produces a different variance at each level for factors measured at that level. Bryk, et. al. (1988) cited four advantages of HLM over regular linear models. First, it can explain student achievement and growth as a function of school-level or classroom-level characteristics while taking into account the variance of student outcomes within schools. Second, it can model the effects of student characteristics, such as gender, race-ethnicity, or socioeconomic status (SES), on achievement within schools or classrooms, and then explain differences in these effects between schools or classrooms using school or classroom characteristics. Third, it can model the between and within-school variance at the same time, and thus produce more accurate estimates of student outcomes. Finally, it can produce better estimates of the predictors of student outcomes within schools and classrooms, by "borrowing" information about these relationships from other schools and classrooms. HLM models are discussed in the literature under a number of different names by different authors from a number of diverse disciplines (Bryk and Raudenbush, 1992; Dempster, Rubin and Tsutakawa, 1981; Elston and Grizzle, 1962; Goldstein, 1987; Henderson, 1984; Laird and Ware, 1982; Longford, 1987; Mason, Wong, and Entwistle, 1984; Rosenberg, 1973).

Extending this methodology to the teacher level becomes more complex. The issue really is not one of whether or not student achievement data should be used in teacher evaluation, but rather entails a methodological debate over ways to operationalize and implement such a system. Unfortunately, the preponderance of literature in the field concentrates upon reasons student achievement data cannot be used for teacher evaluation rather than upon credible ways to use it. Some of the concerns raised in the literature include:

- the development of procedures to account for the difficulty in measuring the long-term development of skills which may not be measured in year-to-year growth patterns (TEA, 1988).

- the assessment of diverse areas of achievement which do not have readily available standardized tests is an area of concern when dealing with non-academic area teachers.

- programs which pull out students for remediation, programs which involve team-teaching, and programs with extensive use of instructional aides inhibit the estimation of an individual teacher's contribution to improved student achievement.

- norm-referenced standardized tests sample broad subject domains and are unlikely to match closely the curriculum in particular classrooms at particular times (Haertel, 1986).

- well-established, broadly applicable, and accepted achievement measures are not available in all the relevant areas of learning (Bano, 1985).

- standardized achievement tests are unlikely to reflect the full range of instructional goals in their subject areas. Norm-referenced tests tend to ignore the higher-order skills. Therefore it is likely that products of superior teaching are not measured adequately or completely by standardized achievement tests (Bano, 1985).

- what the student brings to the classroom in terms of ability, home and peer influence, motivation and other influences is very powerful in affecting academic achievement at the end of the year (Iwanicki, 1986).

- the statistical methods used to control for non-teacher factors cannot take into account all of the relevant factors. More importantly, the methods will be incomprehensible to those being evaluated and difficult to defend in public (Bano, 1985).

- non-statistical models for controlling non-teacher factors are easier to explain, but cannot take into account most of the necessary circumstances (Bano, 1985).

- attempting to use any one of a number of regression-based techniques at the teacher level creates a rather subtle problem related to the statistical concept of "degrees of freedom." In general, the number of degrees of freedom upon which a statistical procedure is based depends on the sample size (N) and the number of sample statistics (i.e., variables in multiple regression). The sample size (i.e., number of students) for a teacher is relatively small to start with. However, the usable sample size becomes even smaller because development of the regression equations requires existing test scores for each student for at least two successive years. As an example, a second-grade teacher may have a class of 22 students, but may only have test scores from the first grade for 11 of those students. Since degrees of freedom also depends on the number of variables in the multiple regression equation, a regression equation with four (4) variables would leave just seven (7) degrees of freedom. The stability of a projected regression line is primarily dependent on the number of degrees of freedom. Seven is generally not enough for stable estimates. As a general rule of thumb, thirty students

per variable has been recommended as a minimum number upon which to base a projected regression line.

Nontechnical concerns most often found in the literature include the concern that objectives that are not measured by the tests will be omitted by teachers, that other duties such as playground supervision and school committee work may be slighted, and that, with each teacher being rated separately, the collegiality necessary to building good instructional teams within a building may be damaged.

Most of the methodological issues raised above can be resolved.(1) Longitudinal growth curves, or alternatively, relationships based upon two or more years of data, can be formulated on important outcome variables. In the case of relationships based upon two years of data, replication is necessary to assure greater reliability. (2) Criterion-referenced tests can be developed and used to assess diverse areas of achievement. (3) In cases where there are pull-out or send-in programs, team teaching, or instructional aides, data can be provided at the team level rather than at the individual teacher level. (4) Measures in addition to norm-referenced tests can be used. (5) Constituents are primarily interested in basic skills. To the extent that measures are needed in music, art, physical education, etc., they can be developed. (6) Criterion-referenced tests can be used to measure higher-order thinking skills. In addition, performance testing can be used as one outcome variable with the outcomes being weighted by the reliability of the instruments. (7) What the student brings to the classroom in terms of background variables can be statistically controlled. These variables typically account for 9-20% of the variance in student achievement (Webster, Mendro, and Almaguer, 1993). (8) It has been the authors' experience that gender, ethnicity, limited English proficiency status, and free-or-reduced-lunch status, plus their interactions, account for most of the variance that can be attributed to background variables. They are easy to explain and defend. (9) Non-statistical models for controlling non-teacher factors are misleading and should not be used (Webster and Edwards, 1993). (10) The degrees of freedom problem is real in that one must worry about the stability of the regression line when it is applied to one teacher. At the teacher level, replication over several years is the best safeguard against erring because of small sample size.

Previous studies conducted in the Dallas Public Schools have demonstrated that equations using school means produce spurious results because they do not take into account the within school variance (Mendro and Webster, 1993); that analysis of unadjusted gain scores produced different results than those produced by regression and HLM Models (r=.73 to .80); that the results produced by gain scores analysis were systematically biased against schools with higher than average Black and poor student populations and in favor of schools with higher than average White, economically advantaged, and Hispanic populations (Webster, et. al., 1995); that reporting of absolute test scores without any additional analysis produced results that were systematically biased against schools with higher than average percent of minority, poor, and Black student populations and in favor of schools with higher than average white and economically advantaged populations and that were very different form those produced by HLM and regression analysis (r=.34 to .60)(Webster, et.al, 1995); and that longitudinal HLM and regression analyses using two years of individual student data for prediction without taking into consideration contextual variables produced results that were somewhat consistent with the HLM

and regression models to be discussed in this paper (r=.89 to .93) but that were systematically biased against schools with higher than average minority, Black, and poor populations and systematically biased in favor of schools with higher than average White, Hispanic, and economically advantaged populations (Mendro, et. al., 1994; Webster and Olson, 1988).

This paper examines the applicability of selected HLM and regression models to the identification of school and teacher effect.

Methodology

Sample

The sample used in this study consisted of all students in the Dallas Public Schools who were in grade 3 in 1994 and grade 4 in 1995 and who had complete data in reading and mathematics. These represent longitudinal cohorts. The temptation to use simulated data was great, however, one of the major purposes of this study was to determine if the HLM routines would execute on real large-scale data sets.

Instrumentation

The instrumentation used for the study was the *Iowa Tests of Basic Skills* Reading and Total Mathematics subtests. Raw scores were the unit of analysis.

Purpose

Five major issues were investigated in this study. They were:

1. What is the correlation among the results produced by the various models for predicting (1) school effect and (2) teacher effect?

2. How much variance is accounted for by each of the alternate models?

3. How consistent are the results?

4. How do the results produced by the various models correlate with individual student and aggregate school demographic variables? (ethnicity, socioeconomic status, English proficiency)

5. How do the results produced by the various models correlate with individual student and aggregate school pre-score characteristics?

Analysis

The analysis consisted of a series of regression and hierarchical linear models. All analyses, except where specified, were completed on residuals that were obtained from solving a series of student level equations designed to account for the effects of ethnicity, limited English proficient status, gender, socioeconomic status, and their first and second order interactions. Equations were developed for both predictor and criterion variables. The unit of analysis for the second-level regression and HLM equations was the residuals obtained from the aforementioned first-level equations.

Specifications for the equations follow.

School Effects

At the school level eleven different models were tried. All models included two stages. The first stage, or fairness stage, was designed to take the effects of important contextual variables out of the subsequent second stage equations for both predictor and criterion variables. Variables used in the first regression and prediction stage included:

Y_{lij} = Outcome variable of interest for each student i in school j. l is a measure for grade/subject/year.
X_{1ij} = Black English Proficient Status (1 if black, 0 otherwise).
X_{2ij} = Hispanic English Proficient Status (1 if Hispanic, 0 otherwise).
X_{3ij} = Limited English Proficient Status (1 if LEP, 0 other).
X_{4ij} = Gender (1 if male, 0 if female).
X_{5ij} = Free or Reduced Lunch Status (1 if subsidized, 0 otherwise).
X_{6ij} = School Mobility Rate (same for all i in each j).
X_{7ij} = School Overcrowdedness (same for all i in each j).
X_{8ij} = Block Average Family Income
X_{9ij} = Block Average Family Education Level
X_{10ij} = Block Average Family Poverty Level
X_{kij} = indicates the variable k for i^{th} student in school j for $i = 1, 2, ..., I_j$ and $j = 1, 2, ..., J$.

The model was

$$Y_l = X\Lambda_l + \varepsilon_l, \qquad \varepsilon_l \sim MVN(\bar{0}, I\sigma^2)$$

and

$$r_l^{95} = Y_l^{95} - \hat{Y}_l^{95}$$
$$r_l^{94} = Y_l^{94} - \hat{Y}_l^{94}$$

Y_l^{94}, Y_l^{95} = Student's scores in 93/94 and 94/95 respectively, for math and reading.

Variables used in the second or prediction stage included:

Student Level Variables:

r_{ij}^{95} = Posttest Residual Score from fairness stage for measure *l* for i^{th} student in school *j*. In this paper it represents *ITBS* Reading 1995 or *ITBS* Mathematics 1995.

r_{hij}^{94} = h^{th} predictor used to estimate r_{ij}^{95} for i^{th} student in school *j*. This is a Pretest Residual score from the fairness stage. In this paper ir represents *ITBS* Reading 1994 and *ITBS* Mathematics 1994.

r_{lij} = $Y_{lij} - \hat{Y}_{lij}$ from OLE

School Level Variables:

W_{1j} = School Mobility
W_{2j} = School Overcrowdedness
W_{3j} = School Average Family Income
W_{4j} = School Average Family Education
W_{5j} = School Average Family Poverty Index
W_{6j} = School Percentage on Free or Reduced Lunch
W_{7j} = School Percentage Minority
W_{8j} = School Percentage Black
W_{9j} = School Percentage Hispanic
W_{10j} = School Percentage Limited English Proficient

Because of limitations in the number of variables that can be used in HLM due to the small n's in some schools, and the high correlations among the background variables, the methodology described in the first stage (fairness stage) was utilized so that HLM was actually run on residuals. The first step in using HLM involves centering the data. The data may be centered around a grand mean or around individual school means. In this study data were centered around the grand mean. If data are not centered, severe problems with multicolinearity are encountered and the HLM program cannot invert many of the matrices associated with individual schools.

School level models included:

DALLAS-FULL (1.0)

Stage 1:[3]

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_6 X_{6ij} + \Lambda_7 X_{7ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

where $\delta_{ij} \sim N(0,\sigma^2)$.

HLM-FULL (2.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_6 X_{6ij} + \Lambda_7 X_{7ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

and

$$\beta_{kj} = \gamma_{k0} + u_{kj}$$

for $i = 1, 2, \ldots, I_j$
$j = 1, 2, \ldots, J$
$k = 0, 1, 2.$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

---

[3] In all regression procedures for both stage 1 or stage 2, arrays were standardized to assure that schools with unusual numbers of students in certain areas of predictor space were not rated based upon differential variance in different arrays.

DALLAS-MC (3.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

where $\delta_{ij} \sim N(0,\sigma^2)$.

HLM-MC (4.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + u_{kj}$$

for $i = 1, 2, \ldots, I_j$
$j = 1, 2, \ldots, J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

DALLAS-MCL (5.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

where $\delta_{ij} \sim N(0,\sigma^2)$.

HLM-MCL (6.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k6} W_{6j} + u_{kj}$$

for $i = 1, 2, \ldots, I_j$
$j = 1, 2, \ldots, J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

HLM-MCC01 (7.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k3} W_{3j} + \gamma_{k4} W_{4j} + \gamma_{k5} W_{5j} + u_{kj}$$

for i = 1, 2, ..., I_j
j = 1, 2, ..., J
k = 0, 1, 2,
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

HLM-MCC02 (8.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k3} W_{3j} + \gamma_{k4} W_{4j} + \gamma_{k5} W_{5j} + \gamma_{k6} W_{6j} + u_{kj}$$

for i = 1, 2, ..., I_j
j = 1, 2, ..., J
k = 0, 1, 2,
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

HLM-MCC03 (9.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_{11}(X_{1ij}X_{4ij}) + \Lambda_{12}(X_{2ij}X_{4ij}) + \Lambda_{13}(X_{3ij}X_{4ij}) + \Lambda_{14}(X_{1ij}X_{5ij}) + \Lambda_{15}(X_{2ij}X_{5ij}) + \Lambda_{16}(X_{3ij}X_{5ij}) + \Lambda_{17}(X_{4ij}X_{5ij}) + \Lambda_{18}(X_{1ij}X_{4ij}X_{5ij}) + \Lambda_{19}(X_{2ij}X_{4ij}X_{5ij}) + \Lambda_{20}(X_{3ij}X_{4ij}X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k3} W_{3j} + \gamma_{k4} W_{4j} + \gamma_{k5} W_{5j} + \gamma_{k6} W_{6j} + \gamma_{k7} W_{7j} + u_{kj}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

HLM-MCC04 (10.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_{11}(X_{1ij}X_{4ij}) + \Lambda_{12}(X_{2ij}X_{4ij}) + \Lambda_{13}(X_{3ij}X_{4ij}) + \Lambda_{14}(X_{1ij}X_{5ij}) + \Lambda_{15}(X_{2ij}X_{5ij}) + \Lambda_{16}(X_{3ij}X_{5ij}) + \Lambda_{17}(X_{4ij}X_{5ij}) + \Lambda_{18}(X_{1ij}X_{4ij}X_{5ij}) + \Lambda_{19}(X_{2ij}X_{4ij}X_{5ij}) + \Lambda_{20}(X_{3ij}X_{4ij}X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k3} W_{3j} + \gamma_{k4} W_{4j} + \gamma_{k5} W_{5j} + \gamma_{k6} W_{6j} + \gamma_{k7} W_{7j} + \gamma_{k8} W_{8j} + \gamma_{k9} W_{9j} + u_{kj}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

HLM-MCC05 (11.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_{11}(X_{1ij}X_{4ij}) + \Lambda_{12}(X_{2ij}X_{4ij}) + \Lambda_{13}(X_{3ij}X_{4ij}) + \Lambda_{14}(X_{1ij}X_{5ij}) + \Lambda_{15}(X_{2ij}X_{5ij}) + \Lambda_{16}(X_{3ij}X_{5ij}) + \Lambda_{17}(X_{4ij}X_{5ij}) + \Lambda_{18}(X_{1ij}X_{4ij}X_{5ij}) + \Lambda_{19}(X_{2ij}X_{4ij}X_{5ij}) + \Lambda_{20}(X_{3ij}X_{4ij}X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k0} + \gamma_{k1}W_{1j} + \gamma_{k2}W_{2j} + \gamma_{k3}W_{3j} + \gamma_{k4}W_{4j} + \gamma_{k5}W_{5j} + \gamma_{k6}W_{6j} + \gamma_{k7}W_{7j} + \gamma_{k8}W_{8j} + \gamma_{k9}W_{9j} + \gamma_{k10}W_{10j} + u_{kj}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

Figure 1 summarizes the various school level models.

Teacher Effects[4]

In attempting to attribute teacher effects seven different models were examined. The question of interest involves the complexity of equations that one must implement in order to produce reliable and valid results. If one could limit the equations to a two-level HLM with the second level being the school level with adjustment for shrinkage to estimate teacher effects one would be able to quality control the system better than if one had to have a different equation or equations for each teacher. With parsimony in mind, the following equations were examined for efficiency in attributing teacher effect.

---

4 Teacher Effectivness Indices are used as part of the needs assessment in the teacher evaluation system. Teachers are not evaluated based on effectiveness indices.

At this point another reference index to student test scores and residuals from stage 1 was added. To wit:

Y_{itj} = i^{th} student with t^{th} teacher in school j,
for i = 1, 2, ..., I_{tj}
for t = 1, 2, ..., T_j
for j = 1, 2, ..., J.

This new index does not affect the model specifications in the previous equations. The student residuals from stage 1 were calculated as $r_{itj} = Y_{itj} - \hat{Y}_{itj}$, for one 95 and two 94 student test scores, and were used at stage 2 to obtain the predicted score, $\hat{r}_{itj}^{95}$ for r_{itj}^{95}.

The residual from stage 2 for i^{th} student for t^{th} teacher from j^{th} school was

$$s_{itj} = r_{itj}^{95} - \hat{r}_{itj}^{95}.$$

s_{itj} was used to calculate the TEIs,
I_{tj} is the number of students for teacher i in school j.

DALLAS FULL & HLM-MCC05 (1.0 and 11.0)

The residuals s_{itj} were aggregated with respect to teachers as follows:

$$\bar{s}_{tj} = \frac{s_{\bullet tj}}{I_{tj}} = \frac{\sum_{i=1}^{I_{tj}} s_{itj}}{I_{tj}}.$$

The TEI for teacher t in school j = $\bar{s}_{tj}\sqrt{I_{tj}}$.

DALLAS FULL & HLM-MCC05 with Shrinkage adjustment.(1.0 and 11.0)

Method A:

Each student residual was treated as an outcome predicting a teacher's performance. Hence each teacher has as many performance indicators as students she/he taught. A student may have counted twice or more for each course a teacher teaches.

Let

$$\mu = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\sum_{i=1}^{I_{tj}} s_{itj}}{\sum_{j=1}^{J}\sum_{t=1}^{T_j} I_{tj}}$$

$$\sigma^2 = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\sum_{i=1}^{I_{tj}} (s_{itj}-\mu)^2}{\sum_{j=1}^{J}\sum_{t=1}^{T_j} I_j}.$$

To calculate the Best Linear Unbiased Prediction (B.L.U.P.) of s_{itj} for the t^{th} teacher in school j,

Let

$$\bar{s}_{tj} = \frac{\sum_{i=1}^{I_{tj}} s_{itj}}{I_{tj}}$$

$$\sigma^2_{tj} = \frac{\sum_{i=1}^{I_{tj}} (s_{itj}-\bar{s}_{tj})^2}{I_{tj}}$$

is the error variance for TEI for teacher t in school j,
then

$$TEI_{tj} = \mu + (\bar{s}_{tj}-\mu)\left(\frac{\sigma^2}{\left(\sigma^2 + \frac{\sigma^2_{tj}}{I_{tj}}\right)}\right)$$

If $\frac{\sigma_{tj}^2}{I_{tj}}$ is large relative to σ^2, the TEI's is biased towards the population mean μ.

If $\frac{\sigma_{tj}^2}{I_{tj}}$ is small relative to σ^2, the TEI's is biased towards the sample mean $\bar{s}_{tj}$ since it is more stable.

Method B:

Each teacher teaches many classes and many courses. The mean student residuals for each course within each class were treated as an outcome predicting a teacher's performance. Thus if a teacher taught two classes, 2A and 2B, the subjects of ITBS MATH and ITBS READING, he/she had 4 performance indicators.

Each s_{itj} was aggregated by class/course and $\bar{s}_{tj}^k$ for k^{th} course/section for teacher t in school j was obtained. There were k_{tj} many of these.

Let

$$\mu = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\sum_{k=1}^{K_{tj}}\bar{s}_{tj}^k}{\sum_{j=1}^{J}\sum_{t=1}^{T_j}K_{tj}}$$

$$\sigma^2 = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\sum_{k=1}^{K_{tj}}(\bar{s}_{tj}^k-\mu)^2}{\sum_{j=1}^{J}\sum_{t=1}^{T_j}K_{tj}}.$$

To calculate the B.L.U.P. of $\bar{s}_{tj}^k$ for the t^{th} teacher in school j,

Let

$$\bar{s}_{tj} = \frac{\sum_{k=1}^{K_{tj}}\bar{s}_{tj}^k}{k_{tj}}$$

$$\sigma^2_{tj} = \frac{\sum_{k=1}^{K_{tj}} (\bar{s}^k_{tj} - \bar{s}_{tj})^2}{k_{tj}}$$

is the error variance for TEI for teacher t in school j,
then

$$TEI_{tj} = \mu + (\bar{s}_{tj} - \mu)\left(\frac{\sigma^2}{\left(\sigma^2 + \frac{\sigma^2_{tj}}{k_{tj}}\right)}\right).$$

TWO LEVEL HLM with TEACHER as 2nd LEVEL (12.0 and 12.1)

Y_{lij} = Outcome variable of interest for i^{th} student from teacher j. l is a measure of grade/subject/year.

$X_{1ij}, X_{2ij}, ..., X_{10ij}$ are the fairness variables for stage 1.

The teacher level variables are :

T_{1ij} = Classroom Percent Mobility
T_{2ij} = Classroom Percent Overcrowdedness
T_{3ij} = Classroom Average Family Income
T_{4ij} = Classroom Average Family Education
T_{5ij} = Classroom Average Family Poverty Index
T_{6ij} = Classroom Percentage on Free or Reduced Lunch
T_{7ij} = Classroom Percentage Minority
T_{8ij} = Classroom Percentage Black
T_{9ij} = Classroom Percentage Hispanic
T_{10ij} = Classroom Percentage Limited English Proficient
T_{kij} = indicates the variable k for i^{th} student in classroom j for $i = 1, 2, ..., I_j$ and $j = 1, 2, ..., J$.

HLM-T (12.0)

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_{11}(X_{1ij}X_{4ij}) + \Lambda_{12}(X_{2ij}X_{4ij}) + \Lambda_{13}(X_{3ij}X_{4ij}) + \Lambda_{14}(X_{1ij}X_{5ij}) + \Lambda_{15}(X_{2ij}X_{5ij}) + \Lambda_{16}(X_{3ij}X_{5ij}) + \Lambda_{17}(X_{4ij}X_{5ij}) + \Lambda_{18}(X_{1ij}X_{4ij}X_{5ij}) + \Lambda_{19}(X_{2ij}X_{4ij}X_{5ij}) + \Lambda_{20}(X_{3ij}X_{4ij}X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r_{ij}^{95} = \beta_{0j} + \beta_{1j} r_{1ij}^{94} + \beta_{2j} r_{2ij}^{94} + \delta_{ij}$$

and

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1}T_{1j} + \gamma_{k2}T_{2j} + \gamma_{k3}T_{3j} + \gamma_{k4}T_{4j} + \gamma_{k5}T_{5j} + \gamma_{k6}T_{6j} + \gamma_{k7}T_{7j} + \gamma_{k8}T_{8j} + \gamma_{k9}T_{9j} + u_{kj}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \tau_{3 \times 3}$, and $\delta_{ij} \perp u_{kj}$.

The TEI for teacher *j* was obtained from the empirical bayes estimate for u_{0j} .

A teacher may have had many TEI's from different subject/courses/classes. These TEI's could be combined directly, or weighted by *n* or combined with a shrinkage adjustment.

HLM-TC (12.1)

Stage 1:

NONE

Stage 2:

$$Y_{ij}^{95} = \beta_{0j} + \beta_{1j} r_{1ij}^{94} + \beta_{2j} r_{2ij}^{94} + \delta_{ij}$$

and

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1}T_{1j} + \gamma_{k2}T_{2j} + \gamma_{k3}T_{3j} + \gamma_{k4}T_{4j} + \gamma_{k5}T_{5j} + \gamma_{k6}T_{6j} + \gamma_{k7}T_{7j} + \gamma_{k8}T_{8j} + \gamma_{k9}T_{9j} + u_{kj}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2,$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \tau_{3 \times 3}$, and $\delta_{ij} \perp u_{kj}$.

A TEI for teacher *j* was obtained from the empirical bayes estimate for u_{0j} .

A teacher may have had many TEI's from different subject/courses/classes. These TEI's could be combined directly, or weighted by *n* or combined with a shrinkage adjustment.

THREE LEVEL HLM MODEL FOR STUDENT/TEACHER/SCHOOL (13.0 and 13.1)

The teacher level variables for school *k* are :

T_{1ijk} = Classroom Percent Mobility
T_{2ijk} = Classroom Percent Overcrowdedness
T_{3ijk} = Classroom Average Family Income
T_{4ijk} = Classroom Average Family Education
T_{5ijk} = Classroom Average Family Poverty Index
T_{6ijk} = Classroom Percentage on Free or Reduced Lunch
T_{7ijk} = Classroom Percentage Minority
T_{8ijk} = Classroom Percentage Black
T_{9ijk} = Classroom Percentage Hispanic
T_{10ijk} = Classroom Percentage Limited English Proficient
T_{pijk} = indicates the variable *p* for i^{th} student from classroom *j* within school *k*, for $i = 1, 2, ..., I_{jk}$, $j = 1, 2, ..., J_k$, and $k = 1, 2, ..., K$.

HLM-3 (13.0)

Stage 1:

$$Y_{ijk} = \Lambda_0 + \Lambda_1 X_{1ijk} + \Lambda_2 X_{2ijk} + \Lambda_3 X_{3ijk} + \Lambda_4 X_{4ijk} + \Lambda_5 X_{5ijk} + \Lambda_{11}(X_{1ijk}X_{4ijk}) + \Lambda_{12}(X_{2ijk}X_{4ijk}) + \Lambda_{13}(X_{3ijk}X_{4ijk}) + \Lambda_{14}(X_{1ijk}X_{5ijk}) + \Lambda_{15}(X_{2ijk}X_{5ijk}) + \Lambda_{16}(X_{3ijk}X_{5ijk}) + \Lambda_{17}(X_{4ijk}X_{5ijk}) + \Lambda_{18}(X_{1ijk}X_{4ijk}X_{5ijk}) + \Lambda_{19}(X_{2ijk}X_{4ijk}X_{5ijk}) + \Lambda_{20}(X_{3ijk}X_{4ijk}X_{5ijk}) + \varepsilon_{ijk}$$

where $\varepsilon_{ijk} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r^{95}_{ijk} = \beta_{0jk} + \beta_{1jk} r^{94}_{1ijk} + \beta_{2jk} r^{94}_{2ijk} + \delta_{ijk}$$

and

$$\beta_{pjk} = \gamma_{p0k} + \gamma_{p1k}T_{1jk} + \gamma_{p2k}T_{2jk} + \gamma_{p3k}T_{3jk} + \gamma_{p4k}T_{4jk} + \gamma_{p5k}T_{5jk} + \gamma_{p6k}T_{6jk} + \gamma_{p7k}T_{7jk} + \gamma_{p8k}T_{8jk} + \gamma_{p9k}T_{9jk} + u_{pjk}$$

for p = 0, 1, 2.

$$\gamma_{pqk} = \alpha_{pq0} + \alpha_{pq1}W_{1k} + \alpha_{pq2}W_{2k} + \ldots + \rho_{pqk}$$

where $E(\delta_{ijk}) = 0$, $Var(\delta_{ijk}) = \sigma^2$, $E(u_{pjk}) = 0$, $Var(u_{pjk}) = \tau_{3\times 3}$, $E(\rho_{pqk}) = 0$, $Var(\rho_{pqk}) = \Sigma$ and $\delta_{ijk} \perp u_{pjk} \perp \rho_{pqk}$.

School Effectiveness Indices were obtained from the empirical bayes residual for ρ_{00k} for school k, $\hat{\rho}_{00k}$.

Teacher Effectiveness Indices for teacher j within school k were obtained from the EB residual for u_{0jk}, $\hat{u}_{0jk}$.

Schoolwide TEIs for teacher j are obtained by combining $\hat{\rho}_{00k} + \hat{u}_{0jk}$.

HLM - 3C (13.1)

Stage 1:

NONE.

Stage 2:

$$Y^{95}_{ijk} = \beta_{0jk} + \beta_{1jk}r^{94}_{1ijk} + \beta_{2jk}r^{94}_{2ijk} + \delta_{ijk}$$

$$\beta_{pjk} = \gamma_{p0k} + \gamma_{p1k}T_{1jk} + \gamma_{p2k}T_{2jk} + \gamma_{p3k}T_{3jk} + \gamma_{p4k}T_{4jk} + \gamma_{p5k}T_{5jk} + \gamma_{p6k}T_{6jk} + \gamma_{p7k}T_{7jk} + \gamma_{p8k}T_{8jk} + \gamma_{p9k}T_{9jk} + u_{pjk}$$

for p=0, 1, 2.

$$\gamma_{pqk} = \alpha_{pq0} + \alpha_{pq1}W_{1k} + \alpha_{pq2}W_{2k} + \ldots + \rho_{pqk}$$

where $E(\delta_{ijk}) = 0$, $Var(\delta_{ijk}) = \sigma^2$, $E(u_{pjk}) = 0$, $Var(u_{pjk}) = \tau_{3\times 3}$, $E(\rho_{pqk}) = 0$, $Var(\rho_{pqk}) = \Sigma$ and $\delta_{ijk} \perp u_{pjk} \perp \rho_{pqk}$.

Figure 2 summarizes the various teacher level models.

BEST COPY AVAILABLE

RESULTS

School Effect

The major objective of this study was to determine an acceptable methodology for identifying how effective various schools and teachers were in addressing the major objectives of schooling. At the school level effect was defined as the difference between a group of students' performance in a particular school and the performance that would have been expected if those students had attended a school with similar context but with practice of average effectiveness. Schools were defined as being above average in effectiveness, at average in effectiveness, or below average in effectiveness. Because the methodology was designed to define effective schools by controlling for factors over which the schools had no control and then to determine which schools made the greatest improvement, the degree of consistency among the results produced by the various least-squares regressions and HLM models was of major interest. While it is obvious that different context and conditioning variables produce different results, it was hypothesized that carefully thought out statistical models utilizing the same context and/or conditioning variables would produce very similar results. Specifically, we were interested in the consistency of results between least-squares regression models that rely on interactions to insure fairness and two-level HLM models that use similar context variables but add conditioning variables at the school level.

Tables 1 and 2 show the correlations between the various models and methods that are specified by the aforementioned school level equations and summarized in Figure 1. The correlations between the results produced by DALLAS-FULL and HLM-FULL, two comparable models, were .9774 in reading and .9633 in mathematics. Similarly, the correlations between the results produced by DALLAS-MC and HLM-MC were .9701 in reading and .9212 in mathematics. Correlations between DALLAS-MCL and HLM-MCL were .9695 and .9119 in reading and mathematics, respectively. Thus the results produced by directly comparable ordinary least squares (OLS) models and HLM models were virtually identical with over 90% of the variance being accounted for. As the models become increasingly different, the correlations drop slightly although all of the correlations in reading are above .91 and all in mathematics are above .86. It seems obvious that the two approaches, one regression-based using first and second order interactions, and one two-level HLM using bayesian adjustments to school level regression lines, produced very similar results. It is also obvious that the eleven different models using OLS or HLM methods and slightly different variables produced very consistent results. Consistency of results is very important since this addresses the reliability of different models for ranking schools We would have liked to enter all of the context variables and their interactions into the first level of HLM to determine if there were differential effects of contextual variables within schools, but with anything but the very simplest of models we couldn't invert the matrices. Thus we were left with the choice of using a contextually rich model that included most of the variables that are significantly related to student achievement, or a more simple model that excluded many of those variables. We chose the approach of having a first regression and prediction stage and computing all of the OLS and HLM models on residuals produced in that first stage.

Remembering that the major objective of this methodology was to assure fairness in comparing the products of schools, one important consideration was that of whether or not individual student background characteristics were related to results. Table 3 presents the correlations between results and various student characteristics. Perusal of Table 3 shows that these correlations were both practically and statistically insignificant. The only student characteristic that was significantly related to outcomes was posttest score, a situation that was expected and desired.

The next important concern was whether or not school level contextual characteristics were related to the results produced by the various models. We know from previous research that school level contextual characteristics such as percentage of low socioeconomic students often correlate with results when individual level contextual characteristics do not (Webster and Olson, 1988). That is, it is often more difficult to move a low socioeconomic student immersed in a school of low socioeconomic students than it is to move a low socioeconomic student enrolled in a school with a number of higher socioeconomic students. Tables 4 and 5 show these correlations in reading and mathematics, respectively. These correlations are neither statistically nor practically significant, meaning that there was no relationship between the results produced by any of these models and the school level variables that were examined. Note that when conditioning variables were introduced at the second level in HLM the correlations with those context variables were adjusted to 0. The student and school level results meant that schools derived no particular advantage from starting with minority or white students, high or low socioeconomic level students, limited or non-limited English proficient students, a high or low mobile student body, or overcrowding or underutilized facilities.

Table 6 displays the correlations of the results provided by the various models and predictor variables (reading residuals 94 and math residuals 94), criterion variables (reading residuals 95 and math residuals 95), and predicted scores. All correlations with predictors were zero, with criteria were significant as expected, and with predicted scores were slightly above zero but statistically and practically non-significant. This means that whether or not a particular student was below, at, or above prediction was not related to the level of the pretest score and therefore that schools derived no particular advantage by starting with high-scoring or low-scoring students.

All things considered, it is important that both student level and school level contextual information be included in models for identifying effective schools. While it may be desirable to include this information in the first level of HLM, the authors were unable to enter sufficient numbers of background variables into the HLM models to reflect the complex nature of these inter-related variables. Rather than oversimplify the models to accommodate a small subset of important context variables within the confines of HLM, a preliminary regression stage was utilized to control for the effects of important context variables. This, in conjunction with a two-stage HLM model, produced minimal correlation between residuals and student level context variables and zero correlation between residuals and school level context variables. Specifically, HLM-MCC05 appears to be the model of choice for determining school effect.

Teacher Effect

Tables 7 and 8 show the correlations between the various models and methods that are specified by the aforementioned teacher level equations and summarized in Figure 2. As can be seen from the data presented in Tables 7 and 8, HLM-MCC05 and HLM-MCC05 B.L.U.P. produced very different results from either the DALLAS-FULL, the STUDENT-TEACHER TWO LEVEL MODEL, or the THREE LEVEL STUDENT-TEACHER-SCHOOL MODEL ($r < .75$). The reason for this is rather straightforward. The HLM-MCC05 models are school level models with the initial student-level equations being calculated within schools. The conditioning variables are school-level conditioning variables that adjust schools' slopes and intercepts for school characteristics. The empirical bayes residuals produced from these equations rank teachers within schools, not across the District. Since we are primarily interested in ranking teachers across the District, not within schools, the HLM-MCC05 models are not appropriate for this purpose. If, however, the school effect is added back into the empirical bayes residuals produced by the HLM-MCC05 equations, the results produced are much more in line with the two level student-teacher and three level student-teacher-school HLM models ($r > .94$). This model is labeled MCC Res+EBb0 and includes the following adjustment to s_{itj}.

$$s_{itj} = r_{itj}^{95} - \hat{r}_{itj}^{95adj}$$

$$\hat{r}_{itj}^{95adj} = \hat{r}_{itj}^{95} + u_{00}^{*}$$

where u_{00}^{*} is the empirical bayes residual for school *j*.

The question then becomes one of which model to use in parsimoniously estimating teacher effect. The advantages of DALLAS FULL, DALLAS FULL-B.L.U.P., HLM-MCC05, HLM-MCC05 B.L.U.P., AND MCC Res+EBb0 are that the equations can be calculated at either the school or the district level, that the number of relevant predictor and conditioning variables is not limited by the methodology as is the case with the three level models, and that all students can be included in the calculations thus allowing most teachers to have indices. The two level student-teacher model and the three-level student-teacher-school model were used as the standard for judging the other models since we believe they produce the best models of teacher effect. We would, however, prefer not to use these models in actual practice since, due to degrees of freedom issues involved with individual teachers, they effectively eliminated from consideration about 20% of teachers who should have had indices.

One interesting factor in these deliberations is that there was very little within teacher variance in the student residuals. This suggest that school effect is really an aggregate teacher effect in that, within schools, there was relatively great between teacher variance in student residuals coupled with little within teacher variance (See Tables 9 and 10). When one examines correlations of results provided by the various models with important teacher level variables (Tables 11 and 12), all correlations except those with class size were statistically and practically non-significant. This means that the various models produced results that are free from bias relative to important classroom level contextual variables. (Class size was not entered into the equations.)

Either of two models produced sufficiently consistent results to be used for estimating teacher effect. DALLAS FULL B.L.U.P. (Least Squares Regression with adjustment for shrinkage) produced results that correlated .9355 and .9140 with the two level student-teacher model and .9120 and .9128 with the three level student-teacher-school model in reading and mathematics, respectively. MCC Res+EBb0 (Two-level student-school HLM with adjustment for shrinkage and with school effect added to the teacher level empirical bayes residuals) produced results that correlated .9684 and .9451 with the two level student-teacher model and .9754 and .9873 with the three level student-teacher-school model in reading and mathematics respectively. Thus we believe that the two level student-school HLM model with adjustment for shrinkage and with school effect added to the teacher level empirical bayes residual produced sufficiently consistent results with those produced by the student-teacher and student teacher-school HLM models to be used in the estimation of teacher effect. The resulting equations can use all available student data and produce indices for the majority of basic skills teachers.

The efficacy of using MCC Res+Ebb0 for determining teacher effect is further supported by an examination of the amount of variance accounted for by each of the models. Table 13 displays the R^2s for each of the models. When examining data, two R^2s are important for each model. The first column for reading and mathematics displays the R^2s from the first fairness stage. In the case of DALLAS-FULL the first stage accounted for 16.96% of the variance in reading. The second stage accounted for an additional 44.67% of the remaining 83.4% of the variance. Thus, between the first and second stage, DALLAS-FULL accounted for 70.75% of the variance in reading. Similar calculations yield the amount of variance accounted for by each of the models. When one examines HLM-MCC05, the base model for MCC Res+EBb0, one determines that one first needs to use the fairness equation from the first stage of DALLAS-FULL and HLM FULL, that is, add average parental income, mobility, and overcrowding back in at the student level. When this is done HLM-MCC05 accounts for over 70% of the variance in both reading and mathematics. This is very close to the variance accounted for by the two level student-teacher HLM equations and the three level student-teacher-school HLM equations.

SUMMARY and DISCUSSION

Several observations appear relevant based on this study. First, and perhaps most important, OLS analysis including first and second order interactions and two-level HLM analysis produced very similar results at both the school (reading, r = .9774; mathematics, r=.9633) and teacher (reading, r = .9530, mathematics, r = .9338) levels. At the teacher level, however, the two-level HLM model with adjustment for shrinkage and with school effect added back into the equations was the model of choice since the results produced by that model correlated very highly with the results produced by two level student-teacher and three level student-teacher-school HLM models (reading, r = .9684 and .9754, respectively; mathematics, r = .9451 and .9873, respectively).

Because the most prevalent method of rating schools is either on absolute test scores or on unadjusted gain scores, it is important to repeat the results of previous studies that demonstrated that such rating systems are biased against schools with higher than average

minority and poor student populations. There are no existing methodological fixes for this short of the use of appropriate statistical models. The fact that the average educator does not comprehend OLS or HLM is no excuse for rating schools or teachers in a haphazard manner that is demonstrably wrong.

Although a previous study (Webster, et.al., 1995) demonstrated that the use of two years of achievement data without contextual variables included in the equations produced results that were different from the results produced by the equations included in this study and that were biased against schools that contained higher than average numbers of Black and economically poor students, we are going to continue investigations in this area. These investigations will include adding a third and fourth year to the prediction and adding contextual variables. Since adding additional areas of matched test scores will significantly reduce the number of students eligible for the analysis, investigations into the use of Bayesian estimation to estimate missing data will also be carried out.

Meanwhile, the models that will be used in Dallas for ranking schools and teachers are as follows.

Y_{lij} = Outcome variable of interest for each student i in school j. l is a measure for grade/subject/year.
X_{1ij} = Black English Proficient Status (1 if black, 0 otherwise).
X_{2ij} = Hispanic English Proficient Status (1 if Hispanic, 0 otherwise).
X_{3ij} = Limited English Proficient Status (1 if LEP, 0 other).
X_{4ij} = Gender (1 if male, 0 if female).
X_{5ij} = Free or Reduced Lunch Status (1 if subsidized, 0 otherwise).
X_{6ij} = School Mobility Rate (same for all i in each j).
X_{7ij} = School Overcrowdedness (same for all i in each j).
X_{8ij} = Block Average Family Income
X_{9ij} = Block Average Family Education Level
X_{10ij} = Block Average Family Poverty Level
X_{kij} = indicates the variable k for i^{th} student in school j for $i = 1, 2, ..., I_j$ and $j = 1, 2, ..., J$.

Student Level Variables:

r_{ij}^{95} = Posttest Residual Score from fairness stage for measure l for i^{th} student in school j. In this paper it represents *ITBS* Reading 1995 or *ITBS* Mathematics 1995.
r_{hij}^{94} = h^{th} predictor used to estimate r_{ij}^{95} for i^{th} student in school j. This is a Pretest Residual score from the fairness stage. In this paper it represents *ITBS* Reading 1994 and *ITBS* Mathematics 1994.
r_{lij} = $Y_{lij} - \hat{Y}_{lij}$ from OLE

School Level Variables:

W_{1j} = School Mobility
W_{2j} = School Overcrowdedness
W_{3j} = School Average Family Income
W_{4j} = School Average Family Education
W_{5j} = School Average Family Poverty Index
W_{6j} = School Percentage on Free or Reduced Lunch
W_{7j} = School Percentage Minority
W_{8j} = School Percentage Black
W_{9j} = School Percentage Hispanic
W_{10j} = School Percentage Limited English Proficient

School Rankings

Stage 1:

$$Y_{ij} = \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_6 X_{6ij} + \Lambda_7 X_{7ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

$$r_l^{95} = Y_l^{95} - \hat{Y}_l^{95}$$
$$r_l^{94} = Y_l^{94} - \hat{Y}_l^{94}$$

Y_l^{94}, Y_l^{95} = Student's scores in 93/94 and 94/95 respectively, for math and reading.

Stage 2:

$$r_{ij}^{95} = \beta_0 + \beta_1 r_{1ij}^{94} + \beta_2 r_{2ij}^{94} + \delta_{ij}$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k3} W_{3j} + \gamma_{k4} W_{4j} + \gamma_{k5} W_{5j} + \gamma_{k6} W_{6j} + \gamma_{k7} W_{7j} + \gamma_{k8} W_{8j} + \gamma_{k9} W_{9j} + \gamma_{k10} W_{10j} + u_{kj}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2.$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

The school rankings are obtained from the empirical bayes residual for β_{00}, which is u^{*}_{00}, where

$$u^{*}_{00} = \beta^{*}_{00} - (\hat{\gamma}_{00} + \sum_{s=1}^{S_q} \hat{\gamma}_{0s} W_{sj})$$

$$\beta^{*}_{00} = \lambda_0 \bar{r}^{95}_{\bullet 0} + (1 - \lambda_0)\hat{\gamma}_{00}$$

$$\lambda_0 = \frac{Var(\beta_{00})}{Var(\bar{r}^{95}_{\bullet 0})}.$$

Teacher rankings

Stage 1:

$$\begin{aligned} Y_{ij} = {} & \Lambda_0 + \Lambda_1 X_{1ij} + \Lambda_2 X_{2ij} + \Lambda_3 X_{3ij} + \Lambda_4 X_{4ij} + \Lambda_5 X_{5ij} + \Lambda_6 X_{6ij} + \Lambda_7 X_{7ij} + \Lambda_8 X_{8ij} + \Lambda_9 X_{9ij} + \\ & \Lambda_{10} X_{10ij} + \Lambda_{11}(X_{1ij} X_{4ij}) + \Lambda_{12}(X_{2ij} X_{4ij}) + \Lambda_{13}(X_{3ij} X_{4ij}) + \Lambda_{14}(X_{1ij} X_{5ij}) + \Lambda_{15}(X_{2ij} X_{5ij}) + \\ & \Lambda_{16}(X_{3ij} X_{5ij}) + \Lambda_{17}(X_{4ij} X_{5ij}) + \Lambda_{18}(X_{1ij} X_{4ij} X_{5ij}) + \Lambda_{19}(X_{2ij} X_{4ij} X_{5ij}) + \Lambda_{20}(X_{3ij} X_{4ij} X_{5ij}) + \varepsilon_{ij} \end{aligned}$$

where $\varepsilon_{ij} \sim$ i.i.d. $\sim N(0,\sigma^2)$.

Stage 2:

$$r^{95}_{ij} = \beta_0 + \beta_1 r^{94}_{1ij} + \beta_2 r^{94}_{2ij} + \delta_{ij}$$

$$\begin{aligned} \beta_{kj} = {} & \gamma_{k0} + \gamma_{k1} W_{1j} + \gamma_{k2} W_{2j} + \gamma_{k3} W_{3j} + \gamma_{k4} W_{4j} + \gamma_{k5} W_{5j} + \gamma_{k6} W_{6j} + \gamma_{k7} W_{7j} + \gamma_{k8} W_{8j} + \gamma_{k9} W_{9j} + \\ & \gamma_{k10} W_{10j} + u_{kj} \end{aligned}$$

for $i = 1, 2, ..., I_j$
$j = 1, 2, ..., J$
$k = 0, 1, 2.$
where $E(\delta_{ij}) = 0$, $Var(\delta_{ij}) = \sigma^2$, $E(u_{kj}) = 0$, $Var(u_{kj}) = \sigma^2$, and $\delta_{ij} \perp u_{kj}$.

$$s_{itj} = r^{95}_{itj} - \hat{r}^{95adj}_{itj}$$

$$\mu = \frac{\sum_{j=1}^{J} \sum_{t=1}^{T_j} \sum_{k=1}^{K_{tj}} \bar{s}^{k}_{tj}}{\sum_{j=1}^{J} \sum_{t=1}^{T_j} K_{tj}}$$

$$\sigma^2 = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\sum_{k=1}^{K_{tj}}(\bar{s}_{tj}^{k} - \mu)^2}{\sum_{j=1}^{J}\sum_{t=1}^{T_j} K_{tj}}.$$

To calculate the B.L.U.P. of $\bar{s}_{tj}^{k}$ for the t^{th} teacher in school j,

Let

$$\bar{s}_{tj} = \frac{\sum_{k=1}^{K_{tj}} \bar{s}_{tj}^{k}}{k_{tj}}$$

$$\sigma_{tj}^2 = \frac{\sum_{k=1}^{K_{tj}}(\bar{s}_{tj}^{k} - \bar{s}_{tj})^2}{k_{tj}}$$

is the error variance for TEI for teacher t in school j,
then

$$TEI_{tj} = \mu + (\bar{s}_{tj} - \mu)\left(\frac{\sigma^2}{\left(\sigma^2 + \frac{\sigma_{tj}^2}{k_{tj}}\right)}\right).$$

References

Aiken, L. S. and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions.* Newburg Park: Sage

Bano, S. M. (1985). *The Logic of Teacher Incentives.* Washington, D.C.: National Association of State Boards of Education.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newburg Pass, California: Sage Publications

Bryk, A. S., Raudenbush, S. W., Seltzer, M., and Congdon, R. (1988) *An Introduction to HLM: Computer Program User's Guide* (2nd ed.) Chicago, Ill: University of Chicago

Cohen, J. (1968). *Multiple Regression As A General Data - Analystic System.* Psycological Bulletin, 70, 426-443.

Cohen, J. and Cohen, P. (1983) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.) Hillsdale, New Jersey: Laurence Erlbaum.

Council of Chief State School Officers (1995). *State Education Accountability Reports and Indicator Reports: Status of Reports Accross the States*, Washington, D. C.

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. V., (1981). *Estimation in Covariance Components Models.* Journal of the American Statistical Association, 76, 341-353.

Darlington, R. B. (1990) *Regression and Linear Models.* New York: McGraw-Hill

Elston, R. C. and Grizzle, J. E. (1962). *Estimation of Time Response Curves and Their Confidence Bands.* Biometrics, 18, 148-159.

Felter, M. and Carlson, D. (1985). Identification of Exemplary Schools on a Large Scale. In Austin and Gerber (eds.), *Research on Exemplary Schools.* New York: Academic Press, 83-96.

Goldstein, H., (1987). *Multilevel Models in Educational and Social Research*. New York: Oxford University Press.

Good, T.L., & Brophy, J.E. (1986). *School effects*. In M.C. Wittrock (Ed.), Handbook of research on teaching (3rd ed.) (pp. 570-602). New York, NY: Macmillan Publishing Company.

Haertel, E. (1986), *One Valid Use of Student Performance Measures for Teacher Evaluation*. Education Evaluation and Policy Analysis, 8, 45-60.

Henderson, C. R., (1984). *Applications of Linear Models in Animal Breeding*. Guelph, Canada: University of Guelph.

Iwanicki, E. G. (1986). *Continuities and discontinuities among measures of teacher performance*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Jaeger, R. M. (1992). Weak Measurement Serving Presumptive Policy. *Kappan*, 74, 2, 118-128.

Kirst, M. (1986). New Directions for State Education Data Systems. *Education and Urban Society*, 18, 2, 343-357.

Klitgaard, R. E. and Hall, G. R. (1973). *A Statistical Search for Unusually Effective Schools*. Santa Monica, Ca.: Rand Corporation.

Laird, N. M. and Ware, H. (1982). *Random-Effects Models for Longitudinal Data*, Biometrics, 38-963-974.

Longford, N. T. (1987). *A Fast Scoring Algoritem for Maximum Likelihood Estimation In Unbalanced Mixed Models with Nested Random Effects.*, Biometrika, 74 (4), 817-827.

McKenzie, D. (1983). School Effectiveness Research: A Synthesis and Assessment. In P. Duttweiler (ed.), Educational Productivity and School Effectiveness. Austin, Texas: Southwest Educational Development t Laboratory.

May, J. (1990). *Real World Considerations in the Development of an Effective School Incentive Program.* ERIC ED 320-271.

Mason, W. M. , Wong, G. Y. and Entwistle, B. (1984). *Contextual Analysis Through the Multilevel Linear Model.* In Leinhardt (ed.) *Sociological Methodology*, 1983-84, San Francisco: Jossey-Bass, pp. 72-103.

Mendro, R. L., Webster, W. J., Bembry, K., and Orsak, T. H. (1994). *An Application of Hierarchical Linear Modeling In Determining School Effectiveness,* Rocky Mountain Educational Research Association, Phoenix, Arizona.

Mendro, R. L., & Webster, W. J. (1993). *Using school effectiveness indices to identify and reward effective schools.* (Paper presented at the Rocky Mountain Research Association, October, 1993, Las Cruces, New Mexico).

Millman, J. (ed.) (1981). *Handbook of Teacher Evaluation.* Beverly Hills, California, Sage.

Rosenberg, B. (1973). *Linear Regression With Randomly Dispersed Parameters* Biometrika, 60, 61-75.

Saka, T. (1984). *Indicators of School Effectiveness: Which are the Most Valid and What Impacts Upon Them*? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA: (ERIC No. ED 306277).

Sanders, W. L., & Horn, S. P. (1995). *The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment, in Shinkinlfred, A. J. and Stufflebeam, D. L.*, Teacher Evaluation: Guide to Effective Practice, Boston: Kluwer.

Texas Education Agency (1988). *A Study to Determine the Most Effective Means of Implementing Career-Ladder Assignments That Are Made on the Basis of Student Achievement In Addition to Other Bases Required By Law.* Austin, Texas: Texas Education Agency, ERIC ED 309 143.

Webster, W. J., Mendro, R. L., and Almaguer, T. (1994). *Effectiveness Indices: A "Value Added" Approach to Measuring School Effect.* Studies in Educational Evaluation. 20, 113-145.

Webster, W. J., Mendro, R. L., Bembry, K., and Orsak, T. H. (1995). *Alternative Methodologies for Identifying Effective Schools*. Distinguished Paper Session, American Educational Research Association, San Francisco, California, ERIC EA 027 189.

Webster, W. J., & Olson, G. H. (1988). *A Quantitative Procedure for the Identification of Effective Schools*. Journal of Experimental Education, 56, 213-219.